# Geospatial Data Management Standards and Procedures
## Supporting the Exchange and Replicating of Geospatial Data

# Appendix B
## VGIS Data Warehouse
## Standards and Procedures for Data Publishers

**Overview:** This document outlines standards and procedures to be used by Publishers*[1] to load and maintain GIS data available via Vermont's GIS (VGIS) Data Warehouse (http://www.vcgi.org/dataware/). VCGI has implemented a *distributed* data distribution architecture that provides users with a single portal for discovering and downloading Vermont's GIS data. This system is referred to as the Distributed VGIS Data Warehouse (Warehouse). The Warehouse consists of several components, including GIS data, a cataloging database, metadata, and ArcIMS data extract services and/or pre-packaged files. The Warehouse is administered by VCGI, but is supported by state agencies and others who create, maintain, and host GIS data for download.

Data "publishers" can link into the system with one of the following two options:

a. **Level I nodes:** VCGI hosts and posts the Publisher's data and metadata. Publisher provides VCGI "snapshots" of their data as defined in the "Geospatial Data Exchange and Replication Protocol". All layers are cataloged within the VGIS Data Warehouse cataloging database.
b. **Level II nodes:** Publisher hosts and posts data and metadata on their own web or FTP server. All layers are cataloged within the VGIS Data Warehouse cataloging database, which points to the Publisher's files (posted on the web/FTP server).

**Resources:** The following resources will help Publishers maintain their VGIS Data Warehouse entries.

- Appendix A – VGIS Themes and Sub-themes = Documents a set of ISO data themes/categories which have been adopted by the VGIS Data Warehouse system.
- Appendix B - Standards and Procedures for Data Publishers = You're looking at it.
- Appendix C –Add/Update Checklist = Checklist used when loading or updating Publisher data available via the Warehouse.
- WAREHOUSE_SUPPORT.gdb = File Geodatabase containing layers and other resources.
- TEMPLATE_LAYER = Template feature class used to ensure consistent XYZ domains, resolution, and tolerances. Stored in WAREHOUSE_SUPPORT.gdb.

**Distributed VGIS Data Warehouse Standards:** GIS data accessible via the VGIS Data Warehouse must meet the following requirements.

1. File Download Format: Zipped (ZIP) ESRI Shapefile (pre-packaged or on-the-fly extract).
2. Metadata: Layers must include FGDC compliant metadata.
3. ArcIMS Extract Services (*optional*): ArcIMS Extract Services support the VGIS Data Warehouse "zip and ship on-the-fly" capability.
4. OGC Web Map Services (*optional*): The VGIS Data Warehouse "Custom Download Tool" requires OGC complaint WMS services.
5. Pre-packaged data (*note: assuming ArcIMS Extract services aren't used):* Zipped shapefile. Must include FGDC metadata in TXT/HTML format.

[1] **Publishers:** Agencies and departments that "publish" geospatial data.

## Procedures (New/Updated Data):

### I. Layer Identification and Review

a. Snag a blank checklist (NewUpdate_Checklist.doc)
b. Identify layer (coverage/shapefile/PGDB/EGDB/FGDB) to load. Enter name on checklist. Circle "Master Format" type.
c. Enter location of master file (ex: J:\gisdata\hazdata\hazsites)
d. Startup ArcCatalog and navigate to directory containing the input layer.
e. Use ArcCatalog to determine properties/features present in the data layer. Mark the feature types on the checklist.

### II. Categorize and Assign Layer Name (*for new data only*)

**a. If new, determine which theme category** the data falls into by reviewing the "Appendix A – VGIS Themes and Sub-themes" list. Enter the full theme name (including sub-category) on the checklist.

**b. If new, assign a layer_name**. Enter the name on the checklist. The following naming conventions are used for feature datasets and feature classes.

Feature Dataset: **<Theme><Theme_sub>*[2]_<NAME>**

For example, the BNDHASH layer has the following *feature dataset* layer name: BoundaryOther_BNDHASH. This naming convention groups data by theme..

Standalone feature classes or the individual *feature classes* within each *feature dataset* have the following naming convention;

Feature Class: **<Theme>_<NAME>_<featuretype>_<subtype>**

For example, the line feature class in BNDHASH is named Boundary_BNDHASH_line. Feature types include poly, line, point, route, anno, and raster. _<subtype> handles cases in which a *dataset* has more than one feature type (ex: more than one route subclass, annotation subclass, annotation arrows, etc.).

Note: Names cannot exceed 32 characters according to ESRI documentation. Therefore <Theme> and <Theme_sub> are limited to 10 characters each. <NAME> can be up to 12 characters, however, the total character length for any feature dataset or feature class name must not exceed 32. This must be taken into consideration

[2] Refer to Appendix A (VGIS Themes and Sub-Themes) for a list of themes and sub-themes.

when assigning a name.It is recommended that the <NAME> be unique database wide. Tiled data sets must include the tile number as part of the <NAME>.

## III. Load data into Enterprise (ArcSDE) or File Geodatabase

a. Load Data Warehouse "Geoprocessing Environment Settings"
   i. Fire up ArcCatalog.
   ii. Add the "ArcToolbox" pane.
   iii. Right-click in the whitespace and select "Save Settings a To Default". This will save your current settings.
   iv. Right-click in the whitespace and select "Load Settings a From File"
      1. Select "dataware_envsettings.xml"

b. **If using Enterprise Geodatabase** (ArcSDE):
   i. From ArcCatalog go to "*GDB_DDev*". Provide username/password. Note: Data gets loaded into GDB_DDev first to ensure success of all steps and data integrity before integrating into GDB_ <Org> (eg: GDB_ANR).

c. **Else if using File Geodatabase** (FGDB):
   i. Create a temporary File Geodatabase (eg: GDB_temp.gdb).
   ii. Drill into newly created FGDB.

d. If the layer is a **Coverage, shapefile, or Geodatabase (file or personal)**
   i. Right-click → New → Feature Dataset.
   ii. Enter the Layer_Name as noted on your checklist sheet. Be sure to use proper casing. Click next.
   iii. In the Coordinate System dialog box pick the "Import" button to import X/Y/Z settings from the template layer. Select WAREHOUSE_SUPPORT.gdb/TEMPLATE_LAYER. Once imported click next.
   iv. Ignore the Vertical Coordinate System dialog box. Click Next.
   v. Import your features into your newly created Feature Dataset (FDS) by right-clicking on the FDS and selecting "Import → Feature Class (single).." (use "multiple" as needed).
      1. Specify an "output feature class" name matching what you put on your checklist sheet. **REMEMBER: Use the feature class name not the feature dataset name (**when loading "arcs" make sure you change the name to "line"**)**.
   vi. In the "Field Map" portion of the dialog box, delete all fields before and including the <covername>_ID (for coverages). Also delete all fields with $ signs. **Note**: Shape and Fid can't be deleted. Click next.
   **vii.** Click Ok.

e. If the data layer is in a **Raster** format choose "Import → Raster Datasets…".
   i. Select the proper "Input Raster"
   ii. Click on the "Environment" button at the bottom on the dialog window, then choose "Raster Storage Settings".
      1. Select the appropriate "Pyramid resampling technique".
      2. Use the defaults for the rest.
   iii. Once imported, rename the master to match the name on your checklist.

f. **Copy data to production geodatabase:** Now that you've loaded the data into a temporary FGDB (eg: GDB_Temp) or ArcSDE (GDB_DDev), you can move it to your production geodatabase. From ArcCatalog right-click→Copy the feature dataset you created, then paste to production geodatabase. **Note:** If the data layer (and associated lookup tables) already exists, archive them first. Once archived you can delete it so that you can copy the new one in.
   i. **If you use ArcSDE**, copy your data to your GDB_<Org> database (eg: GDB_ANR).
   ii. **Else copy it to the File Geodatabase** (FGDB) that used for geodatabase exchange (eg: GDB_ANR.gdb).

## IV. Create/Update Metadata

a. If this is an existing layer, load master metadata TXT file into a text editor (ex: EnvironHazmat_HWFAC.txt) and update it.
b. If it's a new layer, create an FGDC compliant metadata file for it.
   i. Fire up ArcCatalog and browse to the master layer (eg: shapefile, coverage, etc.) you want to document.
   ii. Highlight it and then click on the Metadata Tab.
   iii. Click on the Import button to import the A-metadata_template.txt file.
   iv. Click the Create/Update metadata button.
   v. You can use the Edit metadata option to update the metadata or export to a TXT file and use a text editor.
   vi. When you're done make sure you export to TXT format and save into ./layers_anr/metadata.
c. Generate an HTML and XML files from the master TXT file (NOTE: never edit the HTML file directly. It should always be generated from the master TXT file).
   i. **If you're hosting a Level II site**, use the ./layers_<org>/metadata/A-REFRESHALL_HTML.bat batch file if you want (if this is a new layer you'll need to update the BAT file). ArcCatalog also works. In ArcCatalog use the Export button and export HTML to ./layers_anr/metadata. Repeat for XML. Enter the name for the file (layer_name) you're saving into the filename block then save.
d. Load metadata into the Feature dataset (FDS) that you copied to the production geodatabase. The metadata should NOT be loaded into each individual Feature class.

Select the feature dataset in ArcCatalog and clicking on the Metadata tab. Click on the "Import Metadata" button. Select "FGDC CSGDM (TXT)" under format from the pull-down list. Browse to the file. Deselect "enable automatic update….". Click OK.

V. **Level I sites only:** Submit copy of data to VCGI
   a. Submit copy of data to VCGI in File Geodatabase format.

**VI. Level II sites only:** Pre-packed data as needed
   **a. Pre-packed downloads (in shapefile format) should only be created if it meets one of the followings test:**
      i. You DO NOT have ArcIMS Server. You generally don't need to pre-package data if you use ArcIMS "extract" services.
      ii. Data is very large, which would put too much load on the ArcIMS "on-the-fly" download (which is the default method for the VGIS Data Warehouse).
      iii. Data can't be distributed in Shapefile format for some reason, for example Raster data (ArcInfo GRIDs, TIFF, IMAGE files, etc.).
   b. **If you answer YES to one of the requirements** noted above, then use these procedures to prepackage data.
      i. If the "master" layer isn't a shapefile you'll need to convert it to one. Either way you need to rename the shapefile to match the convention defined above.
      ii. Use the *featureclass* name for the shapefile name. For example, Environ_HWFAC_point.shp for the layer named EnvironHazmat_HWFAC. Store under ./layers_anr/Shapefiles_for_ArcIMS_ZipShip
      iii. Compress the Shapefile and the metadata TXT and HTM files into a ZIP file format. DO NOT CREATE A SELF-EXTRACTING ZIP…JUST A REGULAR ZIP! Use the following naming convention: <*layername*>.*zip*. Example: TransRoad_RDS.zip.
      iv. Copy data to a WWW accessible folder: .\layers_anr\PrePackaged_Shapefiles. If data is tiled create a sub-directory (./layername/) and put compressed file into it.

VII. **Level II sites that use ArcIMS "Extract":** Register with ArcIMS (if layer is new)
   a. If this layer is new (this does NOT include new tiles for tiled data) and therefore hasn't already been registered with ArcIMS, add an entry into the "DATAWARE_<org>" AXL configuration file.
   b. Add an entry for each feature class (ex: polygons, lines, etc.) *Note: Use existing entries as a model, but don't forget to increment the "id=" values*. SAVE.

**VIII. Level II sites that use ArcIMS "Extract":** Refresh ArcIMS services

## IX. Notify VCGI of New/Updated Layer

a. Send an email message to VCGI's DBA at dba@vcgi.org. Include the following information:
   i. Layer_name
   ii. Post_Date (date you posted the data to the ANR GIS data node)
   iii. Last_Update_Date (last time the data was actually updated)
   iv. URL to self-extracting ZIP (if applicable)
   v. URL to metadata
   vi. Tiled data? Yes/No

## X. Confirm Web Access:

a. If this is an update make sure the data and metadata are accessible from http://www.vcgi.org/dataware.
   i. Connect to VCGI's web site.
   ii. Enter the name of the layer as a keyword.
   iii. Verify that the More Info and Download options work.
   iv. **Tiled Data:** Select the download tool to make sure you can select tiled data. Make sure you can actually download the files (test at least two)

b. If this is a new layer you will need to wait until VCGI has added the layer to the VGIS Data Warehouse cataloging database.

## XI. File Checklist

a. If this is a new layer, grab a manila folder and label it with the layer_name. If it exists grab the existing folder. Put the checklist into the folder and file the folder away.